

## Sampling

---

## Sampling

---

- Collecting data from/about a representative portion of a substance or the population.
- The population we want to know about is called the target population.
- As a general rule the larger the sample, the better it is for estimating characteristics of the population.

## Two Overall Types:

---

- Probability sampling
- Non-probability Sampling

## Two Overall Types:

---

- Probability sampling (Representative samples)
  - Probability samples are selected in such a way as to be representative of the population.
  - They provide the most valid or credible results because they reflect the characteristics of the population from which they are selected.
  - There are two types of probability samples
    - Random
    - Stratified

## Two Overall Types:

---

- Non-Probability
  - Less desirable than probability samples.
  - They do not truly represent a population.
  - Why use?
    - A researcher may not be able to obtain a random or stratified sample, or it may be too expensive.
  - Increasing the validity
    - Approximate random selection
    - Eliminating as many sources of bias as possible
  - Types:
    - Quota
    - Purposive (including snowball)
    - Convenience

## Probability Sampling

---

## Random Sampling

- A simple random sample gives each member of the population an equal chance of being chosen.
- **A technicality...**
  - In this class, and in most of biological science, sampling is done *without replacement*.
    - Once selected, a data point (tree, quadrat location, etc.) is removed from further chance of being selected.
    - If the procedures described generate the same data point (tree, quadrat location, whatever), just skip that random number and randomly select another.

## Cluster Sampling

- In cluster sampling the units sampled are chosen in clusters, close to each other.
- Dividing the population into groups
  - These groups are called clusters or blocks.
  - The clusters are randomly selected, and each element in the selected clusters are used.
  - Usually grouped on geographical location.
- Examples are households in the same street, or successive items off a production line.

## Systematic Sampling

- This is random sampling with a system.
- From the sampling frame, a starting point is chosen at random, and thereafter at regular intervals.
- For example, suppose you want to sample 8 houses from a street of 120 houses.
- $120/8=15$ 
  - so every 15th house is chosen after a random starting point between 1 and 15.
  - If the random starting point is 11, then the houses selected are 11, 26, 41, 56, 71, 86, 101, and 116.

## Stratified Sample

- In a stratified sample the sampling frame is divided into non-overlapping groups called strata by characteristics.
  - e.g. geographical areas, age-groups, genders.
- A sample is taken from each stratum, and when this sample is a simple random sample it is referred to as stratified random sampling.

### Choice of Sample Size for each Stratum

- The first step is to find the total number of staff (180) and calculate the percentage in each group.
  - % male, full time =  $(90 / 180) \times 100 = 0.5 \times 100 = 50$
  - % male, part time =  $(18 / 180) \times 100 = 0.1 \times 100 = 10$
  - % female, full time =  $(9 / 180) \times 100 = 0.05 \times 100 = 5$
  - % female, part time =  $(63/180) \times 100 = 0.35 \times 100 = 35$
- This tells us that of our sample of 40,
  - 50% should be male, full time.
  - 10% should be male, part time.
  - 5% should be female, full time.
  - 35% should be female, part time.
  - 50% of 40 is 20.
  - 10% of 40 is 4.
  - 5% of 40 is 2.
  - 35% of 40 is 14.

### Choice of Sample Size for each Stratum

- In general the size of the sample in each stratum is taken in proportion to the size of the stratum. This is called **proportional allocation**.
- Suppose that in a company there are the following staff:
  - male, full time 90
  - male, part time 18
  - female, full time 9
  - female, part time 63
- and we are asked to take a sample of 40 staff, stratified according to the above categories.

## Non-Probability Sampling

### Purposive

- A select group is targeted with sample obtained in non-random way.
- A non-representative subset of some larger population
- Constructed to serve a very specific need or purpose.
  - A researcher may have a specific group in mind, such as high level business executives.
  - The researcher will attempt to zero in on the target group, interviewing whomever is available.

### Purposive

- Snowball sample
  - A subset of a purposive sample
  - Named because one picks up the sample along the way, analogous to a snowball accumulating snow.
  - A snowball sample is achieved by asking a participant to suggest someone else who might be willing or appropriate for the study
  - Snowball samples are particularly useful in hard-to-track populations, such as truants, drug users, etc.

### Quota Sampling

- In quota sampling the selection of the sample is made by the interviewer, who has been given quotas to fill from specified sub-groups of the population. For example, an interviewer may be told to sample 50 females between the age of 45 and 60.
- There are similarities with stratified sampling, but in quota sampling the selection of the sample is non-random.

### Convenience Sample

- The convenience sample is simply one that happens to come your way.
- Sample of available participants, an accidental sample.

